Virology 385 (2009) 28–38

Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/yviro

Differential HIV-1 integration targets more actively transcribed host genes in neonatal than adult blood mononuclear cells

Brian P. Wellensiek [1], Rajesh Ramakrishnan [2], Vasudha Sundaravaradan [3], Roshni Mehta [4], David T. Harris, Nafees Ahmad *

Department of Immunobiology, College of Medicine, University of Arizona, Tucson, AZ 85724, USA

ARTICLE INFO

Article history:
Received 12 June 2008
Returned to author for revision
30 August 2008
Accepted 31 October 2008
Available online 18 December 2008

Keywords:
HIV-1
Integration
Gene expression
Neonatal cells
Cord blood cells

ABSTRACT

We have recently shown an increased HIV-1 replication and gene expression in neonatal (cord) blood mononuclear cells compared with adult cells, which could be due to HIV-1 integration as it targets active host genes. Here we have characterized 468 HIV-1 integration sites within cord and adult blood T-lymphocytes and monocyte-derived macrophages (MDM) from five donors. Several functional classes of genes were identified by gene ontology to be over represented, including genes for cellular components, maintenance of intracellular environment, enzyme regulation, cellular metabolism, catalytic activity and cation transport. Numerous potential transcription factor binding sites at the sites of integration were identified. Furthermore, the genes at the site of integration, transcription factors which potentially bind upstream of the HIV-1 promoter and factors that assist HIV-1 integration were found to be expressed at higher levels in cord than adult cells. Taken together, these results suggest HIV-1 integration occurred in a more actively transcribed genes in neonatal cells compared with adult cells, which may help explain a higher level of HIV-1 gene expression and replication in neonatal compared with adult cells.

Introduction

The majority of HIV-1 infected infants have a higher level of viremia and progress to AIDS faster than infected adults, including their own infected mothers (MaWhinney et al., 1993; Tovo et al., 1992). While the mechanisms behind this difference remain largely unknown, we have shown that HIV-1 replicates more efficiently in neonatal (cord) blood target cells compared with adult cells, which is influenced at the level of HIV-1 gene expression (Sundaravaradan et al., 2006). Several factors, viral and host, could influence this increased HIV-1 gene expression, including the site of HIV-1 integration within infected host cells. Integration of the HIV-1 genome into the host DNA is a critical event in viral lifecycle and could have a profound effect on viral transcription. Recent studies suggest that HIV-1 integrates within active host genes (Schroder et al., 2002), which would place the host transcription machinery in close proximity to the viral genome resulting in increased transcription, virus production and subsequently a high viremia in infected hosts. Since neonatal immune cells are developing cells, HIV-1 may integrate into highly expressed genes in neonatal compared with adult cells, resulting in increased HIV-1 gene expression and replication in neonatal than adult cells (Sundaravaradan et al., 2006).

HIV-1 integration, which was first believed to be a random event (Holmes-Son et al., 2001), has been the focus of several studies. These studies, however, have lead to varying conclusions. It has been shown that HIV-1 integration may be influenced by base composition (Elleder et al., 2002), presence of alphoid (Carteau et al., 1998) or Alu (Mitchell et al., 2004; Stevens and Griffith, 1996) repeats and DNase I hypersensitive sites (Vijaya et al., 1986). However, more recent studies have found preferential integration within actively transcribed genes (Han et al., 2004; Liu et al., 2006; Schroder et al., 2002; Vincent et al., 1990). Furthermore, association of the viral integrase protein with LEDGF/p75 appears to be capable of directing integration into actively transcribed genes (Ciuffi et al., 2005; Shun et al., 2007). In addition, preferential integration of HIV-1 into transcriptionally active genes suggests that such regions may be important for efficient expression of the viral genome (Corbeil et al., 2001). Studies assessing HIV-1 integration and gene expression have reported that differences in HIV-1 expression levels can be due to integration site selection (Jordan et al., 2001). Other studies have been able to link integration site with trends in viremia within infected patients. In patients with low viremia, HIV-1 was found to be integrated predominantly in GC-rich regions; while in patients with high viremia integration of viral

* Corresponding author. Fax: +1 520 626 2100.
E-mail address: nafees@u.arizona.edu (N. Ahmad).
[1] Present address: Biodesign Institute, Arizona State University, Tempe, AZ, USA.
[2] Present address: Department of Molecular Virology and Microbiology, Baylor College of Medicine, Houston, TX, USA.
[3] Present address: Seattle Biomedical Research Institute, Seattle, WA, USA.
[4] Present address: Division of Hematology-Oncology, Department of Medicine, University of California at Los Angeles, CA, USA.

0042-6822/$ – see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.virol.2008.10.052

genomes was in GC-poor regions (Tsyba et al., 2004). These findings suggest that HIV-1 integration is not a purely random event, and integration at different sites within the human genome can affect HIV-1 pathogenesis and disease progression in various hosts, including neonates and adults.

Expression levels of host factors within infected cells may also have a pronounced effect on HIV-1 gene expression. Host factors are involved in several steps of the viral lifecycle, including integration and viral transcription. Several factors, including LEDGF/p75 (Ciuffi et al., 2005; Shun et al., 2007), Ini1 (Sorin et al., 2006) and p300 (Cereseto et al., 2005), assist integration of the HIV-1 genome into the host chromosome, while others such as BAF (Lin and Engelman, 2003) and Rad18 (Mulder et al., 2002) have been implicated in the integration process. Moreover, differential expression levels of these factors between neonates and adults may affect the efficiency and location of HIV-1 integration. In addition, transcription of HIV-1 genes is aided by host transcription factors such as NF-κB, Sp-1, AP-1, and NFAT following successful integration of the viral DNA (Al-Harthi and Roebuck, 1998; Pessler and Cron, 2004). Other transcription factors, such as YY1, have been shown to inhibit HIV-1 replication (Margolis et al., 1994). Binding of these factors in close proximity to the site of HIV-1 integration may assist in recruitment of other essential factors to the integration site. This, and the fact that several transcription factors bind cooperatively, could affect the levels of these factors at the integration site and the efficiency of viral transcription (Bassuk et al., 1997). Expression levels of these host factors involved in integration and transcription in neonates and adults could consequently affect HIV-1 gene expression and replication.

While HIV-1 integration has been the focus of several studies, a majority of these studies have assessed integration within cell lines (Carteau et al., 1998; Schroder et al., 2002; van't Wout et al., 2003), whereas relatively few studies have analyzed integration within primary cells (Barr et al., 2006; Han et al., 2004; Liu et al., 2006). However, no comparative study has been performed that assessed HIV-1 integration and levels of gene expression in neonatal and adult blood T-lymphocytes and monocyte-derived-macrophages (MDM) with respect to differential HIV-1 gene expression and replication in these cell types (Sundaravaradan et al., 2006). To address this, we have used cord blood in place of neonatal blood because cells from cord blood resemble the cells of neonates in both activation levels and memory status (Mo et al., 1998; Ullum et al., 1997), and is also available in larger quantities than neonatal blood. T-lymphocytes and MDM from 5 different donors of cord and adult blood were infected with various strains of HIV-1, and HIV-1 integration sites were characterized by BLAST and BLAT analyses and potential transcription factor binding sites by MatInspector. In addition, the host gene expression profile at the site of integration and potential transcription factors and factors that assist integration was assessed and compared between cells from cord and adult blood.

In this study, we have characterized 468 integration sites from T-lymphocytes and MDM isolated from 5 different donors of cord and adult blood following HIV-1 infection. We show that HIV-1 integration targets mostly different host genes, with a higher percentage of integrations taking place within more transcriptionally active genes in cord T-lymphocytes and MDM compared with adult cells. Using gene ontology hierarchical analysis, several functional classes of genes were found to be over represented in the 468 genes identified at the site of HIV-1 integration. In addition, numerous potential transcription factor binding sites upstream of the HIV-1 promoter at the integration site, as well as factors that are known to be involved in HIV-1 integration, was found to be expressed at higher levels in cord versus adult cells. These findings may explain why HIV-1 replicated and expressed at a higher levels in neonatal (cord) mononuclear cells compared with adult cells (Sundaravaradan et al., 2006).

Results

Determine the sites of HIV-1 Integration in neonatal and adult blood T-lymphocytes and monocyte-derived macrophages

To determine the sites of HIV-1 integration in neonatal and adult target cells, T-lymphocytes and MDM from five different cord and adult blood donors were infected with an X4 and R5 viruses, HIV-1_{LAV} and HIV-1_{BAL} respectively. Chromosomal DNA, which contained the integrated viral genome, was then isolated, purified (Hirt, 1967) 72 h post-infection, and used to perform inverse PCR to amplify the junction between the viral and cellular DNA (Han et al., 2004). The inverse PCR products were cloned and several clones were sequenced, resulting in identification and characterization of 468 HIV-1 integration sites in primary T-lymphocytes and MDM from 5 different donors of cord and adult blood (Table 1). As a control, DNA was also isolated from uninfected (mock) T-lymphocytes and MDM and subjected to inverse PCR, cloning and sequencing. PCR products were mostly not observed in mock samples, however if slight amplification did occur, the integration sites found within mock DNA did not contain the HIV-1 promoter.

Analysis of the HIV-1 integration sites was performed using BLAST (www.ncbi.nlm.nih.gov/BLAST) and BLAT (www.genome.ucsc.edu/cgi-bin/hgBlat), with selection of chromosomal location based on a scored comparison with the human genome. Using these tools a variety of features were characterized about the integration sites, including the host gene present at that site. Integration frequency among the chromosomes was analyzed, and HIV-1 integration was found within every chromosome of the human genome. Several trends were noticed when integration sites were compared in both the cord and adult T-lymphocytes and MDM, as well as to the random integration control (Fig. 1). The frequency of random integration was determined by calculating the percentage of the human genome that each chromosome accounts for, as random integration would take place at that same frequency. We found that chromosome 1 contained the most sites of HIV-1 integration within adult T-lymphocytes. In contrast, chromosome 17 was the most targeted chromosome for HIV-1 integration in cord T-lymphocytes (Fig. 1A). Interestingly, HIV-1 integration appears to be more evenly

Table 1
Number of HIV-1 integration sites within genes of higher expression

Sample #[a]	Donor[b]	Cell type	# integration sites	% genes at integration site expressed at higher levels	*p*-value
1	Adult	Lymphocytes	22	45	0.02
		Macrophages	21	48	0.15
	Cord	Lymphocytes	24	70	0.000007
		Macrophages	27	57	0.008
2	Adult	Lymphocytes	24	32	0.00002
		Macrophages	25	40	0.001
	Cord	Lymphocytes	24	50	1
		Macrophages	28	61	0.0009
3	Adult	Lymphocytes	24	42	0.004
		Macrophages	21	33	0.00003
	Cord	Lymphocytes	25	48	0.15
		Macrophages	20	67	0.00003
4	Adult	Lymphocytes	23	43	0.008
		Macrophages	25	44	0.01
	Cord	Lymphocytes	24	54	0.04
		Macrophages	22	64	0.0002
5	Adult	Lymphocytes	23	35	0.0001
		Macrophages	20	35	0.0001
	Cord	Lymphocytes	23	65	0.0001
		Macrophages	20	50	1

[a] Represents adult and cord blood that was processed at the same time.
[b] The blood for these experiments was obtained from either a healthy adult or from the umbilical cord blood of a newborn.

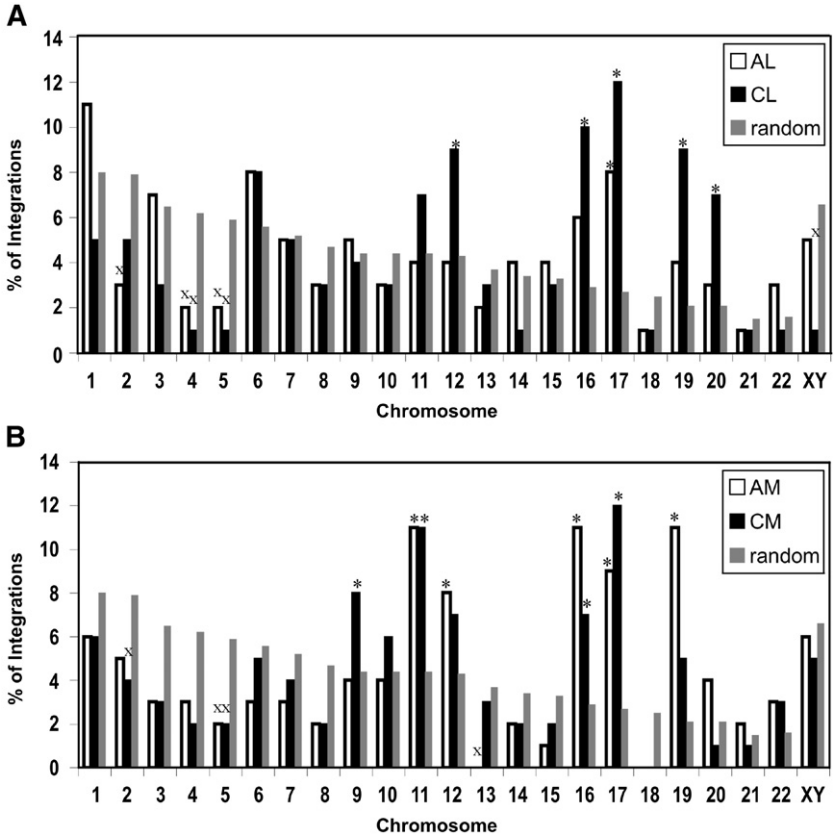

Fig. 1. Chromosomal preferences for HIV-1 integration in T-lymphocytes and MDM. The chromosomal location of the sites of HIV-1 integration were determined by BLAST/ BLAT analysis of the inverse PCR products. The distribution of HIV-1 integration sites within each chromosome was determined for cord and adult T-lymphocytes (A) and MDM (B). The frequency, presented as a percentage, of integration was calculated as the number of integration events in a given chromosome divided by the total number of integration sites found for that sample type. AL: Adult T-lymphocytes; CL: Cord T-lymphocytes; AM: Adult MDM; CM: Cord MDM; Random: Random integration control; *: samples that contain a significantly higher percentage of integration sites compared to random, x: samples that contain a significantly lower percentage of integration sites compared to random.

spread throughout the human genome in adult T-lymphocytes, whereas chromosomes 12, 16, 17 and 19 contain approximately 40% of all integrations in cord T-lymphocytes. Within adult MDM, the most targeted chromosomes were 11, 16 and 19, while chromosomes 11 and 17 frequently contained integration sites in cord MDM (Fig. 1B). Based on these observations, chromosome 11 appears to be a frequent target of HIV-1 integration within MDM, both cord and adult. In addition, chromosome 17 appears to be a frequent target of integration in both cord T-lymphocytes and MDM.

The integration site data was also compared to a random integration control to determine the significance of differential chromosomal targeting by HIV-1 integration. Several chromosomes contained a higher percentage of integration compared to the random control when the integration sites within adult and cord T-lymphocytes was analyzed. For instance in adult T-lymphocytes only chromosome 17 contained a significantly higher percentage ($p<0.02$) while in cord T-lymphocytes, five chromosomes, namely 12, 16, 17, 19 and 20 contained a higher frequency of integration sites ($p<0.03$, 0.007, 0.002, 0.009 and 0.02 respectively) (Fig. 1A). In adult MDM infections chromosomes 11 ($p<0.01$), 12 ($p<0.05$), 16 ($p<0.004$) 17 ($p<0.01$) and 19 ($p<0.002$) and in cord MDM infections chromosomes 9, 11, 16 and 17 ($p<0.05$, 0.01, 0.04 and 0.002 respectively) contained a significantly higher percentage of integration sites (Fig. 1B). When the frequency of integration sites was analyzed as whole for adult and cord T-lymphocytes and MDM it was observed that chromosome 17 contained a significantly higher percentage of integration sites, while chromosome 5 contained a significantly lower percentage of integration in each cell type. In addition, integration frequency was higher in chromosomes 11 and 16 for both adult and cord MDM, and chromosome 4 contained a lower percentage of integration within adult and cord T-lymphocytes.

The difference in integration profiles between adult T-lymphocytes and MDM as well as cord T-lymphocytes and MDM was also analyzed using the data from Fig. 1. When the profiles from adult cells were compared, the greatest difference in the frequency of integration between T-lymphocytes and MDM occurred in chromosome 11, with a 7% higher frequency in MDM. In addition, the frequency of integration within chromosomes 1 and 6 was 5% higher in adult T-lymphocytes, while the frequency in chromosomes 16 and 19 was 5% higher within adult MDM. Analysis of the cord cells profile determined that chromosome 20 displayed a 6% higher frequency of integration within T-lymphocytes compared to MDM, which was the largest difference observed for cord cells. In addition, chromosomes 6 and 19 contained a 4% higher frequency of integration in cord T-lymphocytes, whereas chromosomes 9 and XY displayed a 4% higher frequency of integration within cord MDM.

Several classes of genes were identified as HIV-1 integration targets within both cord and adult T-lymphocytes and MDM by Cytoscape/BiNGO analysis (www.cytoscape.com). Using this method, the gene ontology hierarchy of the host genes found at the site of HIV-1 integration was identified, the genes were grouped based on their function within the host cell, and the biological processes that were significantly over-represented were determined (Fig. 2). The over-represented processes identified in adult T-lymphocytes and MDM were largely related to cellular components and maintenance of the intracellular environment, respectively (Figs. 2A and C). In contrast, genes involved in enzyme regulation and cellular metabolism were over-represented in cord lymphocytes (Fig. 2B), while catalytic activity

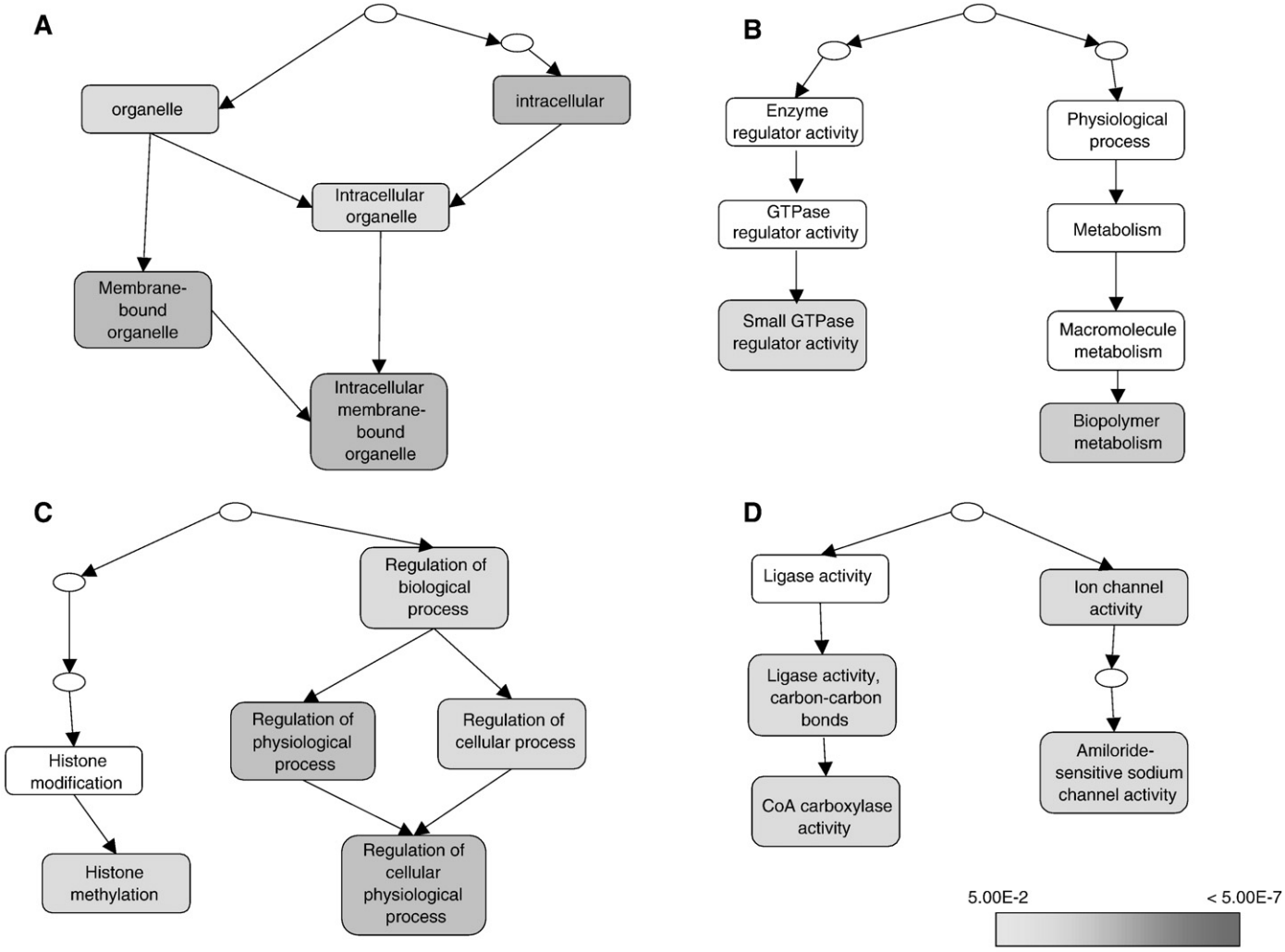

Fig. 2. Over represented classes of genes found at HIV-1 integration site. The host genes present at the sites of HIV-1 integration were determined by BLAST/BLAT analysis of the inverse PCR products. These genes were grouped into classes using a gene ontology hierarchy method, based on cellular function, using BiNGO/Cytoscape. Several classes were found to be statistically over represented. The over represented classes for adult T-lymphocytes (A), cord T-lymphocytes (B), adult MDM (C) and cord MDM (D) are depicted.

and cation transport were over-represented in cord MDM (Fig. 2D). The information on the most specific functional groups from the gene ontology analysis by Cytoscape/BiNGO was combined with the sequence alignment analysis of BLAST/BLAT to specifically characterize the genes in these over-represented functional groups in cord and adult T-lymphocytes (Table 2) and MDM (Table 3).

In addition to classifying genes at the site of HIV-1 integration by function within the host cell, the relevance of these genes to HIV-1 biology was assessed as shown in Table 4. Studies have shown that many genes are upregulated during an HIV-1 infection (van't Wout et al., 2003), and integration into these genes may affect viral transcription and gene expression. The identified genes are involved in various facets of the immune system and the response to viral infection. They are also implicated in several steps of the HIV-1 lifecycle (Table 4).

In our study comparing integration sites in cord and adult cells, there were 10 genes that contained an integration site in both cord and adult T-lymphocytes, and 15 genes in cord and adult MDM. These genes include RAP1A, REPS1, CDK6, SLCO1C1, WNK1, DENND4A, SMURF2, VAV1, NOSIP, TOP1 in T-lymphocytes and EML4, RBMS1, NEK11, HECW1, PLEC1, RXRA, EHMT1, NOTCH1, MS4A1, SLCO1C1, U80760, PDPK1, TMEM49, C20orf14, DMD in MDM. These genes do not appear to be related in function, with several genes performing various functions throughout the cell, and were not believed to be repeated amplification of a single integration event due to the fact that these sites were found by amplification of DNA from different samples. In agreement with other studies (Liu et al., 2006; Schroder et al., 2002), all of the genes which contained integration sites are transcribed by RNA pol II, however the direction of viral transcription did not correlate with the transcriptional direction of the host gene (data not shown). Furthermore, within all genes identified, 95% of all integrations took place within introns, which has been observed in other studies as well (Laufs et al., 2006). This observation is likely to due to the greater sequence length of introns compared to exons (Lander et al., 2001; Venter et al., 2001). In addition, sites of HIV-1 integration were dispersed randomly along the entire length of the genes, showing no bias toward particular positions within the genes. A complete list of all integration sites identified, along with specific information concerning each site is included in the supplemental data.

Identification of potential transcription factor binding sites

Following sequencing of the inverse PCR products, the sequences obtained were analyzed for the identification of potential binding sites for transcriptional factors using the MatInspector program (http://www.genomatix.de/products/ MatInspector/index.html). Many potential transcription factor-binding sites were identified using MatInspector as shown in Table 5. These factors were from a wide variety of transcription factor families and perform very diverse

Table 2
Genes found at sites of HIV-1 integration within T-lymphocytes grouped by functional class

Class	Chromosomal locus	Gene	Gene Description	Dist to start site[a]	I or E[b]	Exp. in adult[c]	Exp. In cord[d]	Ratio[e]
Adult T-lymphocytes								
Intracellular membrane bound organelle	1p36.33	SDF4	Calcium binding protein Cab45	11211	I	4001	20822	+~5.2
	3p21.31	RBM5	RNA binding motif protein 5	9948	I	1113	765	−~1.5
	6p21.31	ITPR3	Inositol 1,4,5-triphosphate receptor type 3	15627	I	604	625	+~1.1
	11q13.1	PACS1	Phosphofurin acidic cluster sorting protein 1	33570	I	1067	1536	+~1.4
	16p13.3	NME4	Nucleoside-diphosphate kinase 4	3237	I	875	1195	+~1.4
	Xq28	IDH3G	Isocitrate dehydrogenase 3 gamma isoform	3171	I	739	941	+~1.3
Cord T-lymphocytes								
Biopolymer metabolism	1p31.3	NFIA	Nuclear factor 1 A-type	203260	I	2248	1831	+~1.2
	6p21.33	VARS	Valyl-tRNA synthetase	12691	I	306	399	+~1.3
	7q21.2	CDK6	Cyclin-dependent kinase 6	23739	I	697	875	+~1.3
	10p15.1	PRKCQ	Protein kinase C, theta	32603	I	1187	1715	+~1.5
	11p13	CSTF3	Cleavage stim. factor subunit 3	11530	I	461	488	+~1.1
	12p13.31	USP5	Ubiquitin specific peptidase 5	11367	E	1189	787	−~1.5
	16q22.1	WWP2	WW domain containing E3 ubiquitin protein ligase	117652	I	1456	1008	−~1.5
	17p13.1	CHD3	Chromodomain helicase DNA binding protein 3	19526	I	919	1193	+~1.3
	17q24.1	SMURF2	SMAD specific E3 ubiquitin protein ligase 2	41765	I	334	420	+~1.3
	17q25.1	CDK3	Cyclin dependent kinase 3	22932	I	291	400	+~1.4
	19p13.3	NFIC	Nuclear factor 1 C-type	101747	I	1831	2248	+~1.2
	20q12	TOP1	DNA topoisomerase 1	5536	I	461	494	+~1.1
	7q31.1	DOCK4	Dedicator of cytokinesis 4	359606	I	338	418	+~1.2
Small GTPase regulator activity	15q26.1	IQGAP1	IQ motif containing GTPase activating protein 1	86003	I	1412	1172	−~1.2
	17q25.3	PSCD1	Cytohesin 1	104192	I	2830	1908	−~1.5
	19p13.3	VAV1	Vav 1 proto-oncogene	14967	I	816	665	−~1.2
	19p13.11	MYO9B	Myosin IXB	27127	I	327	419	+~1.3
	20q13.13	ARFGEF2	ADP ribosylation factor guanine exchange protein 2	107600	I	570	625	+~1.1

[a] The estimated distance, in base pairs, of the HIV-1 integration site from the transcription start site of the host gene.
[b] The presence of the HIV-1 integration site within an intron (I) or exon (E) of the host gene.
[c] Expression of the indicated host gene in adult cells measured in relative intensity.
[d] Expression of the indicated host gene in cord cells measured in relative intensity.
[e] Ratio of host gene expression in cord cells to expression in adult cells. A positive value indicates higher expression in cord cells, a value of 1 indicates even expression and a negative value indicates higher expression in adult cells.

functions within the cell. A number of these identified factors have also been shown to positively impact HIV-1 replication. Several studies have shown that NF-κB, Sp-1, AP-1, and NFAT bind to the viral promoter and increase HIV-1 gene transcription (Al-Harthi and Roebuck, 1998; Pessler and Cron, 2004). Other factors such as YY1 (Margolis et al., 1994), p53 (Duan et al., 1994), c-Myc (Jiang et al., 2007), c-Rel (Doerre et al., 1993), STAT3 (Kohler et al., 2003), STAT5 (Crotti et al., 2007) and E2F (Kundu et al., 1995) have also been found to affect viral replication. In our study, all of these factors were identified as having potential binding sites at the viral integration site,

Table 3
Genes found at sites of HIV-1 integration within MDM grouped by functional class

Class	Chromosomal locus	Gene	Gene Description	Dist to start site[a]	I or E[b]	Exp. In adult[c]	Exp. In cord[d]	Ratio[e]
Adult MDM								
Regulation of cellular physiological processes	2q11.2	EIF5B	Eukaryotic translation initiation factor 5B	38727	I	1066	1443	+~1.4
	12q21.2	SYT1	Synaptotagmin 1	238862	I	1435	2130	+~1.5
	15q15.3	EIF3S1	Eukaryotic translation initiation factor 3	9839	I	2300	1573	−~1.5
	16q22.1	NFATC3	Nuclear factor f activated Tcells, cytoplasmic 3	110355	I	692	841	+~1.2
	17q11.2	TIAF1	TGFB1-induced antiapoptotic factor 1	1325	E	759	651	−~1.2
	17q25.3	FOXK2	Forkhead box K2	40239	I	4152	4601	+~1.1
	19p13.3	ARID3A	AT rich interactive domain containing protein 3A	41574	I	3566	2128	−~1.7
	19p13.3	NFIC	Nuclear factor 1 C-type	56226	I	1831	2248	+~1.2
	19p13.3	THRAP5	Thyroid hormone receptor associated protein 5	20386	I	1992	1540	−~1.3
	Xp11.22	PHF8	PHD finger protein 8	43188	I	1473	1005	−~1.5
Histone methylation	9q34.3	EHMT1	EHMT1 protein	34292	I	442	352	−~1.3
	19p13.2	CARM1	Coactivator –associated arginine methyltransferase 1	40292	I	752	629	−~1.2
Cord MDM								
CoA carboxylase activity	12q24.11	ACACB	Acetyl-Coenzyme A carboxylase beta	130578	I	723	686	−~1.1
	13q32.3	PCCA	Propionyl-Coenzyme A carboxylase alpha	172984	I	365	430	+~1.2
Amiloride-sensitivesodium	16p12.1	SCNN1B	Soduim channel, nonvoltage gated 1	54350	I	668	772	+~1.2
Sensitive sodium channel activity	17q12	ACCN1	Amiloride sensitive cation channel 1	1140197	I	391	439	+~1.1

[a] The estimated distance, in base pairs, of the HIV-1 integration site from the transcription start site of the host gene.
[b] The presence of the HIV-1 integration site within an intron (I) or exon (E) of the host gene.
[c] Expression of the indicated host gene in adult cells measured in relative intensity.
[d] Expression of the indicated host gene in cord cells measured in relative intensity.
[e] Ratio of host gene expression in cord cells to expression in adult cells. A positive value indicates higher expression in cord cells, a value of 1 indicates even expression and a negative value indicates higher expression in adult cells.

Table 4
Genes relevant to HIV-1 biology containing an HIV-1 integration site

Chromosomal locus	Gene	Gene Description	Relevance	Ratio[a]
Adult T-lymphocytes				
3q27.1	PSMD2	Proteasome 26S non-ATPase subunit 2	Degradation of viral protein via proteasome (Dueck and Guatelli, 2007)	−~1.2
6q23.3	IL22RA2	IL-22 receptor alpha-2 chain	Innate host resistance to infection (Misse et al., 2007)	−~1.4
20q12	TOP1	DNA topoisomerase 1	Viral DNA synthesis (Shoya et al., 2003)	+~1.1
Cord T-lymphocytes				
6p21.33	HLA-E	MHC class1 E chain	Presentation of viral peptides to CD8 T-cells	+~1.1
16q22.1	NFAT5	Nuclear factor of activated T-cells 5 isoform C	Viral transcription (Ranjbar et al., 2006)	+~1.3
20q12	TOP1	DNA topoisomerase 1	Viral DNA synthesis (Shoya et al., 2003)	+~1.1
21q22.11	IFNAR1	Interferon-alpha receptor 1 precursor	Suseptibility to infection (Diop et al., 2006)	−~1.5
Adult MDM				
5q13.2	CDK7	Cyclin-dependent kinase 7	Viral transcription (Nekhai et al., 2002)	+~1.3
11q23.1	IL18	Interleukin 18	Innate and adaptive immune response (Torre and Pugliese, 2006)	+~1.1
16q22.1	NFAT5	Nuclear factor of activated T-cells 5 isoform C	Viral transcription (Ranjbar et al., 2006)	+~1.3
Cord MDM				
9q33.2	C5	Complement component 5	Immune response to infection (Kacani et al., 2001)	+~1.2
11p11.2	NUP160	Nucleoporin 160 kDA	PIC translocation through nuclear pore complex (Le Rouzic et al., 2002)	+~1.1
12p13.31	CD4	CD4 antigen precursor	Primary viral receptor on host cell	+~1.1

[a] Ratio of host gene expression in cord cells to expression in adult cells. A positive value indicates higher expression in cord cells, a value of 1 indicates even expression and a negative value indicates higher expression in adult cells.

just upstream of the viral promoter. The distance from the HIV-1 promoter varied, however most factors identified displayed binding sites within 200 base pairs. Several other transcription factors were found to contain binding sites near the viral promoter; however the relevance of these factors to HIV-1 pathogenesis is unclear at this time.

Table 5
Potential transcription factor binding sites near HIV-1 promoter at integration site

Transcription Factor	Dist from HIV-1 promoter (bp)[a]	Exp. in adult[b]	Exp. in cord[c]	Ratio[d]
Adult T-lymphocytes				
YY1	36	3845	2406	−~1.6
NF-κB	97	576	606	+~1.1
c-Myc	136	442	445	~1
STAT3	181	302	399	+~1.3
NFAT	190	295	390	+~1.3
E2F	200	349	423	+~1.2
Adult MDM				
NF-κB	41	576	606	+~1.1
c-Rel	51	720	593	−~1.2
STAT5	52	394	459	+~1.2
YY1	55	3845	2406	−~1.6
NFAT5	85	501	527	+~1.1
p53	128	1028	722	−~1.4
Cord T-lymphocytes				
NFAT	9	295	390	+~1.3
E2F	14	349	423	+~1.2
Sp-1	18	337	432	+~1.3
NF-κB	31	576	606	+~1.1
YY1	44	3845	2406	−~1.6
AP-1	161	330	416	+~1.3
Cord MDM				
NF-κB	24	576	606	+~1.1
STAT3	26	302	399	+~1.3
YY1	44	3845	2406	−~1.6
AP-1	55	330	416	+~1.3
p53	85	1028	722	−~1.4
NFAT	101	295	390	+~1.3

[a] The estimated distance, in base pairs, of the transcription factor binding site upstream of the viral promoter.
[b] Expression of the indicated host gene in adult cells measured in relative intensity.
[c] Expression of the indicated host gene in cord cells measured in relative intensity.
[d] Ratio of host gene expression in cord cells to expression in adult cells. A positive value indicates higher expression in cord cells, a value of 1 indicates even expression and a negative value indicates higher expression in adult cells.

Comparison of gene expression profile in cord and adult cells

The expression profile of host genes within cord and adult cells was examined by microarray analysis. Total RNA from the PBMC of each host was isolated and quantified following stimulation with PHA for 36–48 h and cDNA was generated and labeled with different fluorochromes, with Cy5 coupled to cord cDNA and Cy3 coupled to adult DNA. The labeled cDNAs from both cord and adult were crosslinked to a single 22 K human genes chip. The intensities of each dye were recorded, and further analyzed using GeneSpring®. Values for gene expression were calculated as the ratio of messenger RNA signal for the gene of interest in cord versus adult. Within cord cells, 8013 genes were expressed over 2-fold higher compared with adult samples, while 8028 genes were expressed at least 2-fold higher in adults vs. cord. The remainder of the 22,000 genes represented on the microarray chip was either expressed at similar levels in both cord and adult, or were expressed at levels that resulted in less than a 2-fold difference. Expression levels of several genes found by microarray analysis were validated using real-time PCR (data not shown).

Correlation of HIV-1 integration sites and cellular gene expression

Using the information provided by BLAST and BLAT, as well as the Microarray data, the expression of the 468 identified genes at the sites of HIV-1 integration was determined. To assess the percentage of integration within more highly expressed genes in cord versus adult cells; the percentage was first calculated for each of the five cords and adult cells individually (Table 1). These individual percentages were then averaged to give the overall percentage of integration within highly expressed genes in cord and adult. The standard deviation and p-values were also calculated using the individual sample percentages ($n=5$). The design of our integration experiments allowed us to break down cord and adult cells into more specific cell populations, to separately analyze HIV-1 integration within the two major cell types, T-lymphocytes and MDM, which HIV-1 infects. Within T-lymphocytes integration occurs in a significantly higher percentage of genes that have a higher expression in cord cells compared to adult cells (Fig. 3A). A similar trend of increased host gene expression at the site of HIV-1 integration was seen in cord MDM compared with adult MDM (Fig. 3A).

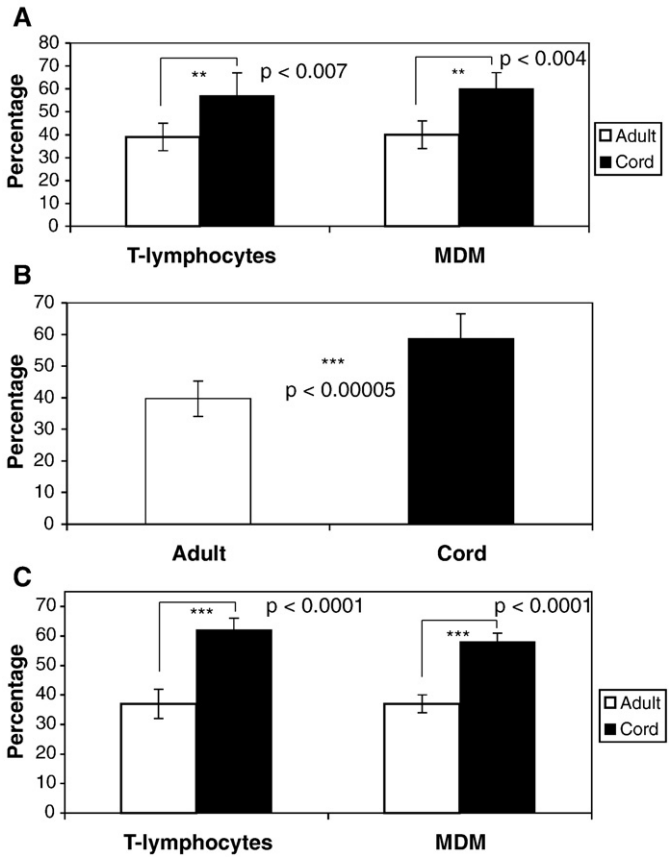

Fig. 3. Host gene expression at the site of HIV-1 integration in cord and adult T-lymphocytes and MDM. The expression of the host gene identified at the site of HIV-1 integration was determined by correlation of data from BLAST/BLAT and microarray analysis. The bars within the graph represent the percentage of genes found at the HIV-1 integration site with higher expression levels in adults and infants (cord) T-lymphocytes and MDM separately (A). Percentages correlating to adult cells are represented by white bars; black bars represent percentages within cord cells. The calculations used to determine the percentages, standard deviations, and p-values are described in the Experimental procedures section. A similar analysis was performed by combining both cell types, T-lymphocytes and MDM, from adults and both cell types from cord to give an overall gene expression profile at the site of HIV-1 integration in adults and infants (cord) (B). In addition, host factors that could potentially bind at the site of HIV-1 integration were determined by analysis of inverse PCR products with MatInspector, which was then correlated with the microarray analysis. The bars within the graph represent the percentage of identified transcription factors, which could possibly bind at the HIV-1 integration site with higher expression levels in adults and infants (cord) T-lymphocytes and MDM separately (C). Percentages correlating to adult cells are represented by white bars; black bars represent percentages within cord cells.

As described above, we also identified HIV-1 integration sites within the same gene in both cord and adult T-lymphocytes and MDM. In these genes, a higher gene expression was observed in both cord T-lymphocytes (70% of genes were higher in cord than adult) and MDM (60% of genes were higher in cord than adult) (data included in Figs. 3A and B). This allowed for a direct comparison of cellular gene expression profile at the same site of HIV-1 integration between cord and adult cells. Furthermore, we determined the overall pattern of host gene expression at HIV-1 integration sites within cord and adult cells and found that a higher number of HIV-1 integration sites were located within more highly expressed genes in cord T-lymphocytes and MDM (239 genes analyzed) compared with adult T-lymphocytes and MDM (229 genes analyzed) (Fig. 3B). In terms of percentage, we found that 59% of all integrations within infected cord cells took place in genes that were expressed at higher levels in cord cells compared to adult cells. In contrast, only 40% of all integrations took place in genes expressed at higher levels in adult cells (Fig. 3B).

Gene expression profile of classes of genes based on gene ontology

Several of the genes identified at the site of HIV-1 integration were grouped into functional classes using gene ontology analysis. In addition, a number of these classes of genes were found to be statistically over represented in the analysis of the integration sites (Fig. 2). By combining the gene ontology data, with the BLAST/BLAT and microarray data, it allowed us to analyze the expression of the genes at the site of HIV-1 integration within over-represented functional gene classes (Tables 1 and 2). This analysis revealed that the genes within most of the classes identified were expressed at higher levels in cord T-lymphocytes (Table 2) and MDM (Table 3) as compared with adult cells. Furthermore, the genes at the site of integration were classified based on involvement in HIV-1 pathogenesis. Most of these genes were also found to be expressed at higher levels in both cord T-lymphocytes and MDM versus adult cells (Table 4).

Gene expression profile of cellular factors associated with HIV-1 biology

In addition to integration site selection, the levels of various host factors could have an impact on HIV-1 gene expression (Kedar et al., 1997). To assess this possibility, the levels of key transcription factors were determined using microarray analysis, and their potential to bind in close proximity to the viral promoter was analyzed using MatInspector. To evaluate the percentage of transcription factors that were expressed at higher levels in cord cells, a statistical analysis was done using similar methods as used with the integration site data ($n=5$). Within T-lymphocytes, a higher percentage of identified transcription factors were expressed at higher levels in cord versus adult cells (Fig. 3C). This trend was also observed in MDM, but to a slightly lesser extent (Fig. 3C). Included within this data are several factors that are known to regulate HIV-1 transcription (Table 5). NF-κB, Sp-1, NFAT, STAT3, STAT5, E2F and AP-1 were expressed at higher levels in cord cells. Furthermore, c-Myc expression was equal in cord compared to adult cells. In addition, factors YY1, c-Rel and p53, which are known to have an inhibitory effects on HIV-1 (Doerre et al., 1993; Duan et al., 1994; Margolis et al., 1994) were found to be expressed at higher levels in adult cells.

Several host factors play pivotal roles in HIV-1 integration, including LEDGF/p75 (Ciuffi et al., 2005; Shun et al., 2007), p300 (Cereseto et al., 2005), BAF (Lin and Engelman, 2003), Rad 18 (Mulder et al., 2002) and Ini1 (Ariumi et al., 2006; Sorin et al., 2006). In our study, all these factors were found to be expressed at higher levels in cord cells compared with adult cells, with the exception of Ini 1. The increased expression of LEDGF/p75 in cord cells is of particular interest, in that this factor has been implicated in directing HIV-1 integration into active genes (Ciuffi et al., 2005; Shun et al., 2007). While these factors do not have as direct an effect on HIV-1 transcription, a majority of these factors were expressed at higher levels in cord versus adult cells, and this increased expression may affect the viral integration process, which in turn may affect viral transcription and gene expression.

Discussion

We have identified 468 HIV-1 integration sites in primary T-lymphocytes and MDM from neonatal (cord) and adult blood of 5 different donors. The genes found at the site of HIV-1 integration were classified into a gene ontology hierarchy, and statistically over represented classes were identified, including genes for cellular components, maintenance of intracellular environment, enzyme regulation, cellular metabolism, catalytic activity and cation transport, as shown in Fig. 2 and Tables 1 and 2. Our data shows that while HIV-1 integration targets actively transcribed genes, as has been seen in other studies (Barr et al., 2006; Han et al., 2004; Levine et al., 2006), integration appears to take place in host genes that are expressed at

higher levels in cord compared with adult cells (Fig. 3). In addition, several potential transcription factor binding sites, just upstream of the viral promoter at the integration site, were identified as shown in Table 5, and these transcription factors were also expressed at higher levels in cord compared to adult cells. Furthermore, the levels of host factors involved in HIV-1 integration, including BAF, p300, Rad18 and LEDGF/p75 (Cereseto et al., 2005; Ciuffi et al., 2005; Lin and Engelman, 2003; Mulder et al., 2002; Shun et al., 2007) were expressed at higher levels in cord cells compared with adult cells. This difference in host gene expression at the site of integration may contribute to the differential HIV-1 gene expression and replication seen in cord blood T-lymphocytes and MDM compared with adult cells (Sundaravaradan et al., 2006), as well as higher viral load and more rapid disease progression in neonates and infants than adults (MaWhinney et al., 1993; Tovo et al., 1992).

Our analysis on HIV-1 integration site distribution throughout the chromosomes revealed several trends. In adult T-lymphocytes, chromosome 1 contained the highest number of integration sites, whereas chromosomes 16, 17, 19, which are gene-dense chromosomes (Lander et al., 2001; Venter et al., 2001), contained a large percentage of integration sites in cord T-lymphocytes. While chromosomes 16 and 19 contained the largest number of integration sites in adult MDM, chromosome 17 had the highest percentage of integration sites in cord MDM. Additionally, in cord and adult MDM chromosome 11, which has been shown to be an integration hot spot (Schroder et al., 2002), contained a high percentage of integration. These results suggest that targeting of HIV-1 integration to gene-dense areas occurs with a greater frequency in cord cells compared to adult cells, which may affect transcription at the site of integration and therefore viral gene expression.

The data concerning integration site selection presented here is also consistent with previous studies (Han et al., 2004; Liu et al., 2006; Schroder et al., 2002; Vincent et al., 1990), however these studies did not analyze integration sites in HIV-1 specific target cells, including T-lymphocytes and MDM, from the same donors. Moreover, comparison of integration site selection in neonatal (cord) and adult target cells that may influence HIV-1 replication (Sundaravaradan et al., 2006) has not been previously explored. Other groups have suggested that integration site selection can lead to different levels of gene expression, and that this may be important for efficient expression of the HIV-1 genome (Jordan et al., 2001). Integration into genes that are actively being transcribed would place the viral genome in relatively close proximity to the transcription machinery (Quivy et al., 2007), thus increasing the efficiency of viral transcription, leading to a higher viral load in infected patients. Our findings within MDM are particularly interesting as we have shown that R5 HIV-1 is transmitted from mother to infant, initially maintained with similar properties (Ahmad et al., 1995; Matala et al., 2001) and replicated efficiently in MDM (Sundaravaradan et al., 2006), which may be critical for the establishment of infection within neonates and infants.

When the overall integration profile was compared for adult and infant (cord) cells, a difference was noticed between the T-lymphocyte profile and the MDM profile. Many differences between the two cell types could account for this interesting trend, one of which could be differences in the nuclear import of the viral pre-integration complex into the host nucleus. While HIV-1 infects both T-lymphocytes and MDM, a difference has been reported in the rate and efficiency of nuclear import between dividing cells, such as T-lymphocytes, and non-dividing cells, such as MDM (Fassati, 2006). Further studies correlating differences in nuclear localization and its effect on integration site selection may provide helpful insights into the exact process by which the HIV-1 genome targets and integrates into the host DNA.

HIV-1 integration site selection was first believed to be a random event (Holmes-Son et al., 2001). Several studies have since uncovered a bias for integration within active genes (Han et al., 2004; Liu et al., 2006; Schroder et al., 2002; Vincent et al., 1990). The study presented here further supports the notion that HIV-1 integration is not random, as the overall integration profile throughout the entire human genome does not coincide with a pattern of random integration (Fig. 1), with several chromosomes containing either a significantly higher or lower percentage of integration sites when compared to a random control. Data from our study has shown several classes of genes within cord and adult T-lymphocytes and MDM which are significantly over-represented, including genes associated with intracellular membrane bound organelles, biopolymer metabolism, histone methylation, and regulation of the cellular environment (Fig. 2, Tables 1 and 2). Further analysis revealed that a majority of the genes in these classes were expressed at higher levels in cord versus adult cells (Fig. 3). Taken together, this information suggests that HIV-1 integration site selection is not random and may result in different outcomes in neonatal and adult target cells. In addition to integration within over-represented functional classes of genes, our study characterized HIV-1 integration into genes that are relevant to HIV-1 biology as shown in Table 4. Several of these genes are known to be involved in HIV-1 pathogenesis and the host response to the viral infection (Diop et al., 2006; Kacani et al., 2001; Misse et al., 2007; Torre and Pugliese, 2006). If HIV-1 integrates into a gene that is highly expressed during a viral infection, it ensures that the viral genome resides within a host gene which is actively transcribed (Schroder et al., 2002). In addition, a majority of these genes were expressed at higher levels in cord compared to adult cells (Table 4). This integration site selection may have an effect on viral gene expression in the very early stages of an infection (Jordan et al., 2001) and differential HIV-1 replication and gene expression in cord versus adult cells (Sundaravaradan et al., 2006), as well as when the infection progresses resulting in a high viremia (Tsyba et al., 2004).

HIV-1 integration takes place when the viral pre-integration-complex (PIC) is formed and migrates to the nucleus of the host cell, where the viral DNA is then irreversibly inserted into the host chromosome (Bukrinsky et al., 1992). This process is facilitated by the viral enzyme integrase and several host factors, including BAF, p300, Rad18, Ini1 and LEDGF/p75 (Cereseto et al., 2005; Ciuffi et al., 2005; Lin and Engelman, 2003; Mulder et al., 2002; Shun et al., 2007; Sorin et al., 2006). Most of these host factors were found to be expressed at higher levels in cord cells compared to adult cells in our study (Fig. 3). An increased expression of these factors could profoundly affect the integration process within these two hosts. Integration within cord cells may, as a result of this expression profile, take place more rapidly and efficiently compared to adult cells. Perhaps increased expression of these factors assists in targeting integration, which has been shown previously with LEDGF/p75 (Ciuffi et al., 2005; Shun et al., 2007), into host genes, and interestingly most of these factors are expressed at higher levels in cord cells compared with adult cells.

Following HIV-1 integration, transcription of the viral genome takes place. This process is solely dependent on host transcription factors at the very early stages of the infection, particularly before the viral transactivator Tat can be produced (Karn, 1999). Results from our study demonstrate that key transcription factors NF-κB, Sp-1, NFAT, STAT3, and AP-1, which have been shown to increase viral transcription (Al-Harthi and Roebuck, 1998; Kohler et al., 2003; Pessler and Cron, 2004), are expressed at higher levels in cord cells compared to adult cells. In addition, the identified factors which are known to repress viral transcription, YY1 (Margolis et al., 1994), p53 (Duan et al., 1994) and c-Rel (Doerre et al., 1993), were expressed at higher levels in adult versus cord cells. Furthermore, these factors have potential binding sites at the site of integration upstream of the viral promoter (Table 5) and could possibly influence viral gene expression and replication in cord versus adult cells. Examples of transcription factors that cooperatively bind to increase transcription have been documented (Bassuk et al., 1997). Additional binding sites for key transcription factors just upstream of the viral promoter, at the site of HIV-1 integration, could therefore possibly enhance recruitment of

the transcription complex or the efficiency with which viral transcription begins.

HIV-1 infected infants have a higher level of viremia and progress more rapidly to AIDS compared with infected adults, including their own infected mothers (MaWhinney et al., 1993; Tovo et al., 1992). However, the mechanisms behind this differential disease progression remain unclear. An explanation, or at least a partial one, has been that an infant's (neonates) immune system is immature and simply is unable to contain the virus (Tiemessen and Kuhn, 2006). The other aspect may be the differential interaction of HIV-1 with neonatal and adult immune cells, which was explored in our previous study which showed that HIV-1 replicated more efficiently in cord target cells compared with adult cells, and was influenced at the level of HIV-1 gene expression (Sundaravaradan et al., 2006). This may result in a high viral load (Abrams et al., 1998; Henrard et al., 1995) and rapid disease progression in neonates and infants (MaWhinney et al., 1993; Tovo et al., 1992). The data presented in our current study shows HIV-1 integration into more highly expressed host genes within cord compared to adult cells may contribute to this increased HIV-1 gene expression and replication in neonatal than adult cells (Sundaravaradan et al., 2006). These results provide novel insights into the mechanisms of differential HIV-1 gene expression, replication and disease progression in neonates and infants compared with adults.

Experimental procedures

Isolation of T-lymphocytes and monocytes/macrophages from cord and adult blood

Cord blood and adult blood from 5 donors, representing various ethnic and racial backgrounds in Arizona, was collected under similar conditions using the same anti-coagulant (heparin sulfate) with written consent and the approval of the University of Arizona Human Subjects Committee. The samples were collected around the same time and processed side by side. Peripheral blood mononuclear cells from the cord blood (CBMC) and adult blood (PBMC) were isolated using a single-step Ficoll-Hypaque method (Amersham-Pharmacia). Monocytes/macrophages (M/M) were separated from the CBMC and PBMC using magnetic cell sorting with anti-CD14 Microbeads (Miltenyi Biotec). T-lymphocytes were collected from the column flow-through, and cultured in RPMI 1640 media containing 10% FBS, penicillin-streptomycin and PHA (2 μg/mL) for 36–48 h. The T-lymphocytes were then washed and cultured in RPMI 1640 media containing 10% FBS, IL-2 (10 U/mL) and penicillin-streptomycin. The M/M from the column were differentiated into macrophages by culturing in RPMI 1640 media containing 15% human AB serum, penicillin-streptomycin and MCSF (100 U/mL) for 7 days. Both cell types were plated in 48-well plates at a density of 1×10^6 cells/well.

Infection with HIV-1

The viruses used in this study were obtained from the National Institutes of Health AIDS Research and Reference Reagent Program. Following culture, the T-lymphocytes and macrophages were infected with the lab-adapted stains HIV-1_{BaL} or HIV-1_{LAV}. Briefly, both cell types were washed and placed in serum free media. Equal amounts of virus were allowed to adsorb to the cells for 2 h at 37 °C. Following adsorption, the serum free media along with any unadsorbed virus was removed and replaced with the appropriate culture media. The infected cells were cultured for 72–84 h.

Lysis and Isolation of DNA

To isolate chromosomal DNA, the HIRT method of DNA extraction was used (Hirt, 1967). Briefly, approximately 1×10^6 cells were centrifuged at 17,000 rpm for 3 min and the cell pellet was resuspended in HIRT solution I (5 mM Tris pH 7.7; 10 mM EDTA), followed by Proteinase K (10 mg/mL) and HIRT solution II (10 mM EDTA; 5 mM Tris pH 7.7; 1.2% SDS). The samples were incubated at 37 °C for 30 min, followed by addition of 4 M NaCl and incubation at 4 °C for a minimum of 3 h. After incubation the samples were centrifuged at 17,000 ×*g* for 45 min at 4 °C and the supernatant was removed. The resulting cell pellet was resuspended in water, and exposed to RNAse (100 mg/mL) for 30 min at 37 °C. Further protein removal was accomplished using phenol-chloroform-isoamyl alcohol (IAA) extraction. The DNA was precipitated and resuspended in sterile, de-ionized water at approximately 10,000 cell equivalents/μl.

Inverse PCR and sequencing

To determine the host chromosomal sequence that flanks the HIV-1 integration site we used inverse PCR (Han et al., 2004). This method uses the restriction enzyme *PstI* to cut the integrated HIV-1 DNA as well as the host chromosome. PstI cuts frequently in genomic DNA however only cuts viral DNA once within the *gag* region. Digested DNA was self-ligated under dilute conditions into circular DNA using T4 DNA ligase. The circularized DNA was used as a template for outer PCR using outward facing primers that are located within the HIV-1 LTR and *gag* regions (LTR out (-): 5′-TAACCAGAGAGACCCAGTA CAG GC-3′; *gag* out (+): 5′-GGTCAGCCAAAATTACCCTATAGTG-3′). This amplification captured the junction between the 5′ end of the viral genome and host chromosomal DNA. The outer PCR product was used for nested PCR to further amplify the viral-host DNA junction (LTR inner(-): 5′-TGGTACTAGCTTGAAGCACCTCCA-3′; *gag* inr (+): 5′-TGTTAAAAGAGAC-CATCAATGAGGAAG-3′). The cycles for both outer and nested PCR were as follows: 94 °C for 3 min, then 30 cycles of 94 °C for 30 s, 65 °C for 1 min, and 68 °C for 2 min, followed by a final extension step of 68 °C for 4 min. The nested PCR product was cloned using TOPO TA cloning and several clones from each cell type were sequenced using the ABI Prism® 3700 DNA automated sequencing system with the M13 reverse primer based vector. As a control, DNA was also isolated from uninfected (mock) T-lymphocytes and MDM and subjected inverse PCR. This allowed for evaluation of nonspecific amplification or amplification due to human endogenous retroviral (HERV) sequences. PCR products were generally not observed in mock samples, however if slight amplification did occur the integration sites did not contain HIV-1 LTR sequences.

Determine sites of integration by BLAST and BLAT analyses

Once the sequences were obtained they were aligned with the HIV-1 LTR to confirm sequencing of the host DNA that flanks the viral promoter. The sequences were then input into BLAST (www.ncbi.nlm.nih.gov/BLAST) and BLAT (www.genome.ucsc.edu/cgi-bin/hgBlat) to determine the host chromosomal gene into which HIV-1 integrated. Both programs contain similar functions and produce similar, but not identical results. These programs perform nucleotide sequence alignments using sequence databases and calculate the statistical significance of matches. Each match is given a numerical score based on the exactness of the alignment and the probability of finding a similar match by random chance, given as an '*E*' value ranging from 1 to 0. Sequence matches were considered acceptable if they resulted in a high numerical score, a low '*E*' value. Several sequences, which contained the HIV-1 LTR, did not match with any known genes in the human genome. These sequences were not included in any further analyses, as this study focused specifically on the expression of genes into which HIV-1 integrated.

Determine functional classes of genes found at HIV-1 integration site

The genes identified at the site of HIV-1 integration by BLAST/BLAT analysis were classified by cellular function using a gene ontology

hierarchy analysis. This analysis was performed using the BiNGO plug-in (www.psb.ugent.be/cbd/papers/ BiNGO) (Maere et al., 2005) in the Cytoscape program (www.cytoscape.com) (Shannon et al., 2003). This analysis identified which gene ontology categories were statistically overrepresented in our set of genes. The gene list was analyzed in BiNGO using a hypergeometric statistical test with Benjamini & Hochberg False Discovery Rate correction at a p-value cut-off of 0.05.

Determine possible transcription factor binding sites by MatInspector

The inverse PCR sequences were also analyzed using MatInspector (http://www.genomatix.de/products/MatInspector/index.html) to assess any possible transcription factor binding sites within the chromosomal region that flanks the HIV-1 integration site (Cartharius et al., 2005). MatInspector uses the information of core positions, nucleotide distribution matrix and a conservation index vector to scan sequences of unlimited length for pattern matches. If the core value associated with a match was above .8, and the matrix similarity value was greater than the defined optimized matrix similarity, then it was considered a more significant match than by random chance, and therefore an acceptable match between the query sequence and the transcription factor binding site sequence. If a section of a submitted sequence resulted in an acceptable match, that position on the sequence was identified as having a potential transcription factor-binding site in that region. Although the transcription factor binding site prediction can infer the binding potential, it cannot infer the functionality of a site.

RNA isolation

Total cellular RNA was isolated from cord and adult blood mononuclear cells, after the cells were stimulated with PHA for 36 to 48 h, using the Qiagen RNeasy kit (Qiagen) following the manufacturer's protocol. UV spectrophotometry was used to quantify the RNA, which was then precipitated in absolute ethanol with 1/10 volume 3 M Sodium Acetate. After centrifugation, the RNA pellet was washed with 70% ethanol made in DEPC water, and resuspended in RNase free water such that the RNA will be present in the concentration needed for microarray analysis.

Microarray

The total RNA was first tested using the Agilent ® BioAnalyzer to ensure the RNA stability, using 18S and 28S peaks as indicators. The differential gene expression in cord and adult cells was measured using a two-color fluorescence hybridization scheme. Each of the samples to be compared was hybridized onto one chip. The labeling, hybridization and acquisition of images was done at the core facility of the Arizona Research Laboratories at the University of Arizona. A pre-printed in-house 22 K human genes chip was used to perform the microarray. In addition to 22 K human genes, each chip included housekeeping genes and non-specific controls. cDNA was generated from the RNA samples provided using random hexamer primers and SuperScript III RT (Invitrogen). Aminoallyl dNTPs were used to label the cDNA, which was cleaned using a QIAquick column (Qiagen). Fluorochromes were conjugated to the cDNA, with Cy3 attached to adult samples and Cy5 attached to cord samples, using an indirect coupling reaction. The DNA on the chips was cross-linked to the hybridization slides in a Stratalinker at 2500×100 μJ. The array was scanned at two wavelengths following independent excitation with an Array WoRx Scanner (Applied Precision Inc). Fluorescent intensity data following excitation was recorded using the SoftWorx™ tracker program, and analyzed using the GeneSpring program (Silicon Genetics Inc.).

Statistical analysis

To assess the percentage of integration within more highly expressed genes and the percentage of transcription factors with higher expression in cord versus adult; the percentage was first calculated for each cord and adult T-lymphocyte and MDM sample individually, yielding five individual percentages. These individual percentages were then averaged to give the overall percentage of integration within highly expressed genes or of transcription factors with higher expression in cord and adult. The standard deviation and p-values were calculated using Student's *t*-test with the individual sample percentages ($n=5$). The significance of frequent integration into particular chromosomes compared to the random integration control was determined using χ^2-tests. In all evaluations, a p value of <0.05 was considered significant.

Acknowledgments

We thank Susan Miller with the University of Arizona Biotechnology Computing Facility for her assistance with the microarray analysis, Naina Bagrodia and Rukmalee Vithana for their assistance with cloning and sequencing, and the AIDS Reference and Reagent Program for providing the HIV-1 isolates used in this study. The author's declare that they have no competing interests. This work was supported by a grant to NA from the Arizona Biomedical Research Commission (7002).

References

Abrams, E.J., Weedon, J., Steketee, R.W., et al., 1998. Association of human immunodeficiency virus (HIV) load early in life with disease progression among HIV-infected infants. New York City Perinatal HIV Transmission Collaborative Study Group. J. Infect. Dis. 178 (1), 101–108.

Ahmad, N., Baroudy, B.M., Baker, R.C., et al., 1995. Genetic analysis of human immunodeficiency virus type 1 envelope V3 region isolates from mothers and infants after perinatal transmission. J. Virol. 69 (2), 1001–1012.

Al-Harthi, L., Roebuck, K.A., 1998. Human immunodeficiency virus type-1 transcription: role of the 5′-untranslated leader region (review). Int. J. Mol. Med. 1 (5), 875–881.

Barr, S.D., Ciuffi, A., Leipzig, J., et al., 2006. HIV integration site selection: targeting in macrophages and the effects of different routes of viral entry. Molec. Ther. 14 (2), 218–225.

Bassuk, A.G., Anandappa, R.T., Leiden, J.M., 1997. Physical interactions between Ets and NF-kappaB/NFAT proteins play an important role in their cooperative activation of the human immunodeficiency virus enhancer in T cells. J. Virol. 71 (5), 3563–3573.

Bukrinsky, M.I., Sharova, N., Dempsey, M.P., et al., 1992. Active nuclear import of human immunodeficiency virus type 1 preintegration complexes. Proc. Natl. Acad. Sci. U. S. A. 89 (14), 6580–6584.

Carteau, S., Hoffmann, C., Bushman, F., 1998. Chromosome structure and human immunodeficiency virus type 1 cDNA integration: centromeric alphoid repeats are a disfavored target. J. Virol. 72 (5), 4005–4014.

Cartharius, K., Frech, K., Grote, K., et al., 2005. MatInspector and beyond: promoter analysis based on transcription factor binding sites. Bioinformatics 21 (13), 2933–2942.

Cereseto, A., Manganaro, L., Gutierrez, M.I., et al., 2005. Acetylation of HIV-1 integrase by p300 regulates viral integration. EMBO J. 24 (17), 3070–3081.

Ciuffi, A., Llano, M., Poeschla, E., et al., 2005. A role for LEDGF/p75 in targeting HIV DNA integration. Nat. Med. 11 (12), 1287–1289.

Corbeil, J., Sheeter, D., Genini, D., et al., 2001. Temporal gene regulation during HIV-1 infection of human CD4+ T cells. Genome Res. 11 (7), 1198–1204.

Crotti, A., Lusic, M., Lupo, R., et al., 2007. Naturally occurring C-terminally truncated STAT5 is a negative regulator of HIV-1 expression. Blood 109 (12), 5380–5389.

Diop, G., Hirtzig, T., Do, H., et al., 2006. Exhaustive genotyping of the interferon alpha receptor 1 (IFNAR1) gene and association of an IFNAR1 protein variant with AIDS progression or susceptibility to HIV-1 infection in a French AIDS cohort. Biomed. Pharmacother. 60 (9), 569–577.

Doerre, S., Sista, P., Sun, S.C., et al., 1993. The c-rel protooncogene product represses NF-kappa B p65-mediated transcriptional activation of the long terminal repeat of type 1 human immunodeficiency virus. Proc. Natl. Acad. Sci. U. S. A. 90 (3), 1023–1027.

Duan, L., Ozaki, I., Oakes, J.W., et al., 1994. The tumor suppressor protein p53 strongly alters human immunodeficiency virus type 1 replication. J. Virol. 68 (7), 4302–4313.

Dueck, M., Guatelli, J., 2007. Evidence against a direct antiviral activity of the proteasome during the early steps of HIV-1 replication. Virology 361 (1), 1–8.

Elleder, D., Pavlicek, A., Paces, J., et al., 2002. Preferential integration of human immunodeficiency virus type 1 into genes, cytogenetic R bands and GC-rich DNA regions: insight from the human genome sequence. FEBS Lett. 517 (1–3), 285–286.

Fassati, A., 2006. HIV infection of non-dividing cells: a divisive problem. Retrovirology 3, 74.

Han, Y., Lassen, K., Monie, D., et al., 2004. Resting CD4+ T cells from human immunodeficiency virus type 1 (HIV-1)-infected individuals carry integrated HIV-1 genomes within actively transcribed host genes. J. Virol. 78 (12), 6122–6133.
Henrard, D.R., Phillips, J.F., Muenz, L.R., et al., 1995. Natural history of HIV-1 cell-free viremia. JAMA 274 (7), 554–558.
Hirt, B., 1967. Selective extraction of polyoma DNA from infected mouse cell cultures. J. Mol. Biol. 26 (2), 365–369.
Holmes-Son, M.L., Appa, R.S., Chow, S.A., 2001. Molecular genetics and target site specificity of retroviral integration. Adv. Genet. 43, 33–69.
Jiang, G., Espeseth, A., Hazuda, D.J., et al., 2007. c-Myc and Sp1 contribute to proviral latency by recruiting HDAC1 to the human immunodeficiency virus type 1 promoter. J. Virol. 81 (20), 10914–10923.
Jordan, A., Defechereux, P., Verdin, E., 2001. The site of HIV-1 integration in the human genome determines basal transcriptional activity and response to Tat transactivation. EMBO J. 20 (7), 1726–1738.
Kacani, L., Stoiber, H., Speth, C., et al., 2001. Complement-dependent control of viral dynamics in pathogenesis of human immunodeficiency virus and simian immunodeficiency virus infection. Mol. Immunol. 38 (2–3), 241–247.
Karn, J., 1999. Tackling Tat. J. Mol. Biol. 293 (2), 235–254.
Kedar, P.S., Arden, K., Foyle, M., et al., 1997. Umbilical cord blood mononuclear cell HIV-1 LTR binding activities. J. Biomed. Sci. 4 (5), 217–228.
Kohler, J.J., Tuttle, D.L., Coberley, C.R., et al., 2003. Human immunodeficiency virus type 1 (HIV-1) induces activation of multiple STATs in CD4+ cells of lymphocyte or monocyte/macrophage lineages. J. Leukoc. Biol. 73 (3), 407–416.
Kundu, M., Srinivasan, A., Pomerantz, R.J., et al., 1995. Evidence that a cell cycle regulator, E2F1, down-regulates transcriptional activity of the human immunodeficiency virus type 1 promoter. J. Virol. 69 (11), 6940–6946.
Lander, E.S., Linton, L.M., Birren, B., et al., 2001. Initial sequencing and analysis of the human genome. Nature 409 (6822), 860–921.
Laufs, S., Guenechea, G., Gonzalez-Murillo, A., et al., 2006. Lentiviral vector integration sites in human NOD/SCID repopulating cells. J. Gene Med. 8 (10), 1197–1207.
Le Rouzic, E., Mousnier, A., Rustum, C., et al., 2002. Docking of HIV-1 Vpr to the nuclear envelope is mediated by the interaction with the nucleoporin hCG1. J. Biol. Chem. 277 (47), 45091–45098.
Levine, B.L., Humeau, L.M., Boyer, J., et al., 2006. Gene transfer in humans using a conditionally replicating lentiviral vector. Proc. Natl. Acad. Sci. U. S. A. 103 (46), 17372–17377.
Lin, C.W., Engelman, A., 2003. The barrier-to-autointegration factor is a component of functional human immunodeficiency virus type 1 preintegration complexes. J. Virol. 77 (8), 5030–5036.
Liu, H., Dow, E.C., Arora, R., et al., 2006. Integration of human immunodeficiency virus type 1 in untreated infection occurs preferentially within genes. J. Virol. 80 (15), 7765–7768.
Maere, S., Heymans, K., Kuiper, M., 2005. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics 21 (16), 3448–3449.
Margolis, D.M., Somasundaran, M., Green, M.R., 1994. Human transcription factor YY1 represses human immunodeficiency virus type 1 transcription and virion production. J. Virol. 68 (2), 905–910.
Matala, E., Hahn, T., Yedavalli, V.R., et al., 2001. Biological characterization of HIV type 1 envelope V3 regions from mothers and infants associated with perinatal transmission. AIDS Res. Hum. Retrovir. 17 (18), 1725–1735.
MaWhinney, S., Pagano, M., Thomas, P., 1993. Age at AIDS diagnosis for children with perinatally acquired HIV. J. Acquir. Immune Defic. Syndr. 6 (10), 1139–1144.
Misse, D., Yssel, H., Trabattoni, D., et al., 2007. IL-22 participates in an innate anti-HIV-1 host-resistance network through acute-phase protein induction. J. Immunol. 178 (1), 407–415.
Mitchell, R.S., Beitzel, B.F., Schroder, A.R., et al., 2004. Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. PLoS Biol. 2 (8), E234.
Mo, H., Monard, S., Pollack, H., et al., 1998. Expression patterns of the HIV type 1 coreceptors CCR5 and CXCR4 on CD4+ T cells and monocytes from cord and adult blood. AIDS Res. Hum. Retrovir. 14 (7), 607–617.
Mulder, L.C., Chakrabarti, L.A., Muesing, M.A., 2002. Interaction of HIV-1 integrase with DNA repair protein hRad18. J. Biol. Chem. 277 (30), 27489–27493.
Nekhai, S., Zhou, M., Fernandez, A., et al., 2002. HIV-1 Tat-associated RNA polymerase C-terminal domain kinase, CDK2, phosphorylates CDK7 and stimulates Tat-mediated transcription. Biochem. J. 364 (Pt 3), 649–657.
Pessler, F., Cron, R.Q., 2004. Reciprocal regulation of the nuclear factor of activated T cells and HIV-1. Genes Immun. 5 (3), 158–167.
Quivy, V., De Walque, S., Van Lint, C., 2007. Chromatin-associated regulation of HIV-1 transcription: implications for the development of therapeutic strategies. Sub-cell. Biochem. 41, 371–396.
Ranjbar, S., Rajsbaum, R., Goldfeld, A.E., 2006. Transactivator of transcription from HIV type 1 subtype E selectively inhibits TNF gene expression via interference with chromatin remodeling of the TNF locus. J. Immunol. 176 (7), 4182–4190.
Schroder, A.R., Shinn, P., Chen, H., et al., 2002. HIV-1 integration in the human genome favors active genes and local hotspots. Cell 110 (4), 521–529.
Shannon, P., Markiel, A., Ozier, O., et al., 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13 (11), 2498–2504.
Shoya, Y., Tokunaga, K., Sawa, H., et al., 2003. Human topoisomerase I promotes HIV-1 proviral DNA synthesis: implications for the species specificity and cellular tropism of HIV-1 infection. Proc. Natl. Acad. Sci. U. S. A. 100 (14), 8442–8447.
Shun, M.C., Raghavendra, N.K., Vandegraaff, N., et al., 2007. LEDGF/p75 functions downstream from preintegration complex formation to effect gene-specific HIV-1 integration. Genes Dev. 21 (14), 1767–1778.
Sorin, M., Yung, E., Wu, X., et al., 2006. HIV-1 replication in cell lines harboring INI1/hSNF5 mutations. Retrovirology 3, 56.
Stevens, S.W., Griffith, J.D., 1996. Sequence analysis of the human DNA flanking sites of human immunodeficiency virus type 1 integration. J. Virol. 70 (9), 6459–6462.
Sundaravaradan, V., Saxena, S.K., Ramakrishnan, R., et al., 2006. Differential HIV-1 replication in neonatal and adult blood mononuclear cells is influenced at the level of HIV-1 gene expression. Proc. Natl. Acad. Sci. U. S. A. 103 (31), 11701–11706.
Tiemessen, C.T., Kuhn, L., 2006. Immune pathogenesis of pediatric HIV-1 infection. Curr. HIV/AIDS Rep. 3 (1), 13–19.
Torre, D., Pugliese, A., 2006. Interleukin-18: a proinflammatory cytokine in HIV-1 infection. Curr. HIV Res. 4 (4), 423–430.
Tovo, P.A., de Martino, M., Gabiano, C., et al., 1992. Prognostic factors and survival in children with perinatal HIV-1 infection. The Italian Register for HIV Infections in Children. Lancet 339 (8804), 1249–1253.
Tsyba, L., Rynditch, A.V., Boeri, E., et al., 2004. Distribution of HIV-1 in the genomes of AIDS patients. Cell. Mol. Life Sci. 61 (6), 721–726.
Ullum, H., Lepri, A.C., Victor, J., et al., 1997. Increased losses of CD4+CD45RA+ cells in late stages of HIV infection is related to increased risk of death: evidence from a cohort of 347 HIV-infected individuals. Aids 11 (12), 1479–1485.
van't Wout, A.B., Lehrman, G.K., Mikheeva, S.A., et al., 2003. Cellular gene expression upon human immunodeficiency virus type 1 infection of CD4(+)-T-cell lines. J. Virol. 77 (2), 1392–1402.
Venter, J.C., Adams, M.D., Myers, E.W., et al., 2001. The sequence of the human genome. Science 291 (5507), 1304–1351.
Vijaya, S., Steffen, D.L., Robinson, H.L., 1986. Acceptor sites for retroviral integrations map near DNase I-hypersensitive sites in chromatin. J. Virol. 60 (2), 683–692.
Vincent, K.A., York-Higgins, D., Quiroga, M., et al., 1990. Host sequences flanking the HIV provirus. Nucleic Acids Res. 18 (20), 6045–6047.